

## University of Groningen

### Hierarchical process modeling

Klein Entink, R.H.; Fox, J.P.; Betlem, B.H.L.; Roffel, B.

*Published in:*  
Journal of Process Control

*DOI:*  
[10.1016/j.jprocont.2006.10.003](https://doi.org/10.1016/j.jprocont.2006.10.003)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2007

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Klein Entink, R. H., Fox, J. P., Betlem, B. H. L., & Roffel, B. (2007). Hierarchical process modeling: Describing within-run and between-run variations. *Journal of Process Control*, 17(4), 349-361.  
<https://doi.org/10.1016/j.jprocont.2006.10.003>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Hierarchical process modeling: Describing within-run and between-run variations

R.H. Klein Entink <sup>a,\*</sup>, J.-P. Fox <sup>a</sup>, B.H.L. Betlem <sup>b</sup>, B. Roffel <sup>b</sup>

<sup>a</sup> University of Twente, Department of Behavioural Sciences, P.O. Box 217, 7500AE Enschede, The Netherlands

<sup>b</sup> University of Twente, Department of Science and Technology, P.O. Box 217, 7500AE Enschede, The Netherlands

Received 3 October 2005; received in revised form 5 July 2006; accepted 18 October 2006

---

## Abstract

Run-to-run variability is a common problem for modeling batch-wise and semi-continuous operated processes. Although observed reactor runs show the same trends in process behaviour, each specific reactor run also shows its own characteristics. Until now, available modeling methods were unable to describe the observed between run variance. In this paper, we present a hierarchical modeling method to solve this problem. A case study for a semi-continuous operated polymer process is analysed to illustrate the hierarchical modeling approach. It was shown that, using the applied modeling method, it is possible to obtain a model which is robust over several reactor runs and which provides a tool for process analysis. Also, the hierarchical model was shown to be superior to a model which lacked an appropriate description of the differences between reactor runs.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Between-run variance; Multi-stage sampling; Hierarchical modeling

---

## 1. Introduction

Polyolefins are used in a wide variety of applications. For example, polypropylene fibers are used for in carpet as well as car parts. These different applications of the same polymer require different polymer properties. Therefore, for most polyolefins, multiple polymer quality grades are produced to be able to specify the different properties. Manufacturers are driven by market demands to produce stringent polymer qualities and minimize their operational costs. For that reason, quality control is vital in polymer production [1,2]. Unfortunately, polymer properties such as density and molecular weight are difficult to measure

online. As a result, polymer quality measurements are available infrequently, which complicates polymer quality control. Conversely, process variables such as flows, pressures, temperature and concentrations are easily and frequently measured. Therefore, an interesting area of research for the polymer industry is the development of mathematical models, which draw inferences about polymer quality from these process measurements [3].

A good review of current modeling techniques is given by Kiparissides [1]. Models can be categorized into two groups:

- phenomenological (white box) models, based on first-principles studies of the process
- empirical (black box) models, based on laboratory experiments or operational data

For a good understanding of a polymer process, a first-principles modeling approach should be followed. However, such an approach is often complicated by lack of

---

\* Corresponding author. Tel.: +31 53 4893650.

E-mail address: [r.h.kleinentink@gw.utwente.nl](mailto:r.h.kleinentink@gw.utwente.nl) (R.H. Klein Entink).

understanding of the process, due to the complexity of the reaction mechanisms of polymerization. This is one of the reasons that black box techniques are often applied. For this reason, a statistical approach is developed in this study for process monitoring purposes.

For our investigation, we show results of a real industrial data example below. We collected process data of 116 reactor runs of two polymer quality grades from a semi-continuously operated polyolefin process. The data include polymer quality measurements, as well as process variables as flows, temperatures and pressures. Although observed reactor runs show the same trends in process behaviour, each specific reactor run also shows its own characteristics: run-to-run variability is observed. Terwiesch et al. [4,5] recognized run-to-run variations in their survey, which were probably due to (unmeasured) impurities or up-stream disturbances. An analogy of run-to-run variability can be seen in the behavioural sciences, for example, in pupil performance studies. In the analysis of pupils' abilities, usually performances of pupils, nested within schools, are studied. Then, school-to-school variability is observed in the pupils' performances, which is often the result of e.g., a different teacher or variations in class or school-specific background variables. Longford and Aitkin [6] recognized the homogeneity between pupils within the same school, sharing a common educational environment, compared to pupils from different schools. Further, they argued that the stratified sampling of pupils nested within schools and the analysis should accordingly recognize and model variance components at each sampling level, that is, the level of individuals and the level of schools.

The fact that measurements are obtained from different reactor runs complicates in two ways a straightforward statistical modeling approach. First, the process circumstances can vary across reactor runs and therefore influential process variables may have different effects on the polymer quality within each reactor run. A proper statistical model that specifies a relation between the measured polymer qualities and various process variables should take into account that the relationships may vary across reactor runs. Second, polymer quality measurements obtained from a specific reactor run are more alike than polymer quality measurements from different reactor runs. A set of reactor run measurements share the same process conditions and for that reason they will presumably have correlated errors. For that reason the observed data of polymer quality measurements of all reactor runs are not independently distributed. In general, the polymer quality measurements within a reactor run are more homogeneous than those of a random sample of measurements from a set of observations from several reactor runs. This greater homogeneity is naturally modeled by a positive within-reactor run correlation among measurements obtained in the same run. This leads formally to a multilevel model where one variance component represents random sampling error

and where another variance component represents a positive covariance structure between measurements from the same reactor run.

Common statistical techniques used to monitor batches are, amongst others, principal component analysis (PCA), partial least squares (PLS) and their multi-way counterparts [7–9]. In PCA and PLS a new set of variables is defined through the projection of the variables onto new orthogonal subspaces. The new variables are a linear combination of the original variables. The statistical models are developed from past batch runs that are “in control” and show similar behaviour. That is, the batches should operate in reactors of similar design, with the same catalysts and the same operational program etc (see, e.g., Nomikos and MacGregor [10]). The process can subsequently be monitored by comparing the process variable trajectories against the “optimal trajectories”. As long as the batches show similar behaviour and similar run-to-run variation, this is a valid approach. Process dynamics can be accounted for by incorporating lagged variables in the analysis. In that case the principal components capture the variables that show the largest variability in a steady state situation as well as dynamic situations. As Kassidas et al. [11] pointed out, this is a valid approach as long as dynamic relationships remain the same. When batches have different length, however, it may be expected that the dynamic relationships between variables is not the same from run-to-run. In addition, models capable of dealing with batches of different length suffer from poor statistics [12]. To deal with this situation Kassidas et al. proposed to use Dynamic Time warping (DTW), a method originating from the area of speech recognition. DTW is a pattern matching method that can expand and compress patterns (scaling) such that similar patterns are matched. In DTW, every process variable receives a certain weight and every warped batch is connected to a reference batch. Kassidas uses the batch which has a run length that is closest to the average run length of all batches as the reference batch. Ramaker et al. [12] suggest to choose the batch which gives the best result in terms of process monitoring as the reference batch. The result of the application of the DTW algorithm is the calculation of an optimal warped time profile that every “normal” batch should follow.

It will be shown that multilevel models are well suited to analyse data from both batch and semi-continuous processes. In comparison to the mentioned traditional statistical techniques, a multilevel model can handle varying effects of process variables across reactor runs and the reactor runs are allowed to be of different lengths without the assumption that all batches show similar behaviour. The analysis of hierarchically structured data (e.g., polymer quality measurements are nested within reactor runs) from the process under consideration shows the need of multilevel models that can handle correlations among observations and allow for varying effects of important process variables across reactor runs. It can be expected that the

polymer quality measurements are influenced by reactor run specific characteristics and by characteristics operating at a different process level over reactor runs. In our real industrial data example, polymer quality measurements are influenced by accumulation of fouling over reactor runs but also by specific concentration values of the catalyst which are controlled by an operator. Note that collective statistical judgments can be made about relationships between the observed polymer quality measurements and relevant process variables, across as well as within reactor runs, due to a multilevel modeling approach. In the proposed approach all observed data are modeled simultaneously such that information can be used from other reactor runs in the estimation of reactor run specific parameters. As a result, stable and accurate parameter estimates can be obtained even for a small number of measurements per reactor run.

## 2. The process

In this study, data are collected from a polyolefin production process. Examples of polyolefin processes are polyethylene, poly(1-butene) and polypropylene production. Different polymer grades are produced for most polyolefins, for general descriptions of such processes see [13]. The structure of the polymer produced is influenced by the addition of a comonomer in varying concentrations, usually of the order of only a few weight per cent. By addition of the comonomer, the reaction mechanism is influenced, since the comonomer also acts as a chain transfer agent. Higher relative concentrations of the comonomer therefore influence the polymer structure, and thereby the properties of the polymer are altered.

### 2.1. Process description

The process studied is a copolymerization, during which a few different polymer qualities (grades) are produced. Polymer quality setpoints are controlled with the feed flow ratio of the two monomers. The monomer feed streams are mixed and, just before entering the reactor, the catalyst is added. The reaction is carried out semi-continuously in a stirred vessel. The process operates at such a temperature, that polymerization is favored over chain transfer, since the latter has a relatively higher activation energy. Due to the exothermic nature of the reaction, cooling is provided by a cooling liquid in the reactor jacket. Fouling of the reactor vessel leads to a decrease of heat transfer, due to which the reaction must be stopped after a certain run length for reasons of safety and production capacity. After reaction, the product stream is fed to a separation unit to obtain the polymer product. The remaining raw materials are recycled to a separation train to remove poisons, chain transfer agents and other by-products from the reaction. The major time constants of the process are the reactor residence time and the residence time of the separation unit.

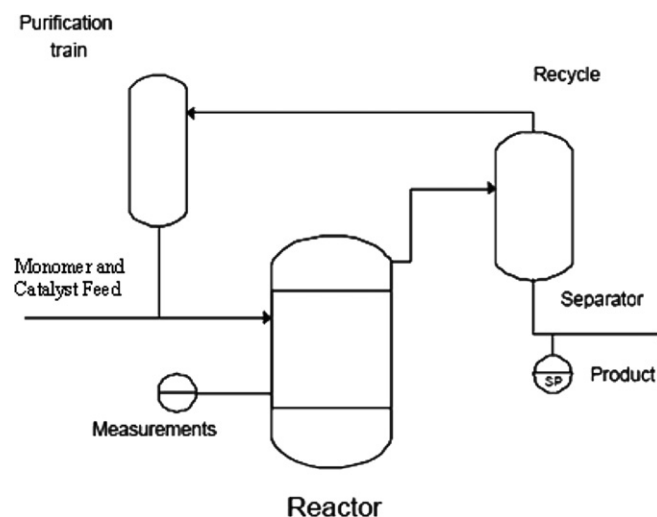


Fig. 1. Schematic process flow diagram.

At the plant, four identical reactors are available for production. A schematic representation of the process is shown in Fig. 1. The main control variables are the feed flow rate and the catalyst to feed flow ratio.

### 2.2. Data collection

Polymer quality samples were available every hour. Measurements during the first hour were excluded so to ignore start-up variation. The reactor runs were of different lengths. However, each reactor run produced a minimum of 10 measurements, since estimates of within-run sampling variance become unstable for shorter runs. Besides the polymer quality indicated by  $Y$ , measured process variables are reactor slurry temperature (at 4 points), catalyst concentration, pure parts catalyst, monomer concentrations at the reactor feed, monomer A conversion, monomer B conversion, catalyst to feed flow ratio, feed flow rate, catalyst efficiency, reactor pressure, stirrer power, reactor heat release and poisons concentration. These variables are indicated by  $x_1, \dots, x_Q$  where  $Q = 13$ . Further, for each process variable the mean value of the reactor run measurements are stored in a variable indicated by  $z_k$  ( $k = 1, \dots, K$ ) that represent process information across reactor runs.

Data were collected from two polymer grades, namely grade A and grade B. For grade A, data of 76 reactor runs from three reactors were collected; in total 966 data points for every variable were collected. For grade B, 40 runs from two reactors were available, with 501 data points for every variable. Each of the four reactors can produce polymer grades A and B. The data were collected by a Dutch company and for reasons of competitiveness specific numerical details are not given. However, this does not complicate the illustration of the multilevel modeling approach towards batch data.

### 2.3. Synchronization of the measurements

Every hour a polymer sample is taken at the outlet of the separation unit. All other process measurements are made upstream the sample point with a sampling time of one minute. The residence times of the reactor and the separation unit thereby introduce large time delays. In order to relate these measurements to the corresponding polymer quality measurement, a dynamic filter is applied. Assuming the reactor and the separation vessel are ideally stirred tanks and assuming isothermal conditions, Roffel and Betlem [14] derived that the response function for each variable  $x_q$ , ( $q = 1, \dots, Q$ ), as a function of the flow rate  $F$  is a constant  $C$ , but with a first order time delay. In the Laplace domain, the response function is given by:

$$\frac{\delta x_q}{\delta F} = C \frac{1}{1 + \phi s}. \quad (1)$$

This means that the measurements of the process variables should be corrected for the residence times of the reactor and the separation vessel, according to a first order filter as (1) which means multiplying with  $\frac{1}{1 + \phi s}$ . Written in exponential form the filtering equation is:

$$x_{q,\text{filtered}}^{t+1} = x_q^t \cdot e^{-1/\phi_1} + (1 - e^{-1/\phi_1}) \cdot x_q^{t+1}, \quad (2)$$

where  $\phi_1$  corresponds to the residence time of the reactor and  $t$  denotes the time step. After correcting for the reactor residence time, the same procedure is applied for the residence time of the separation vessel. At the plant, the residence times were experimentally determined as a function of the feed flow rate. This way a relationship was specified between the process measurements and the corresponding polymer quality sample. A similar derivation for the time response behaviour of ideally stirred tanks can be found in Westerterp, Van Swaaij and Beenackers [15].

## 3. Development of the multilevel model

In this section, a short introduction on the theory of hierarchical modeling will be presented, as far as it is relevant to this investigation.

### 3.1. Design

In statistics, a common assumption is that measurements are randomly drawn samples from an infinite population, drawn independently from each other. However, the semi-continuous operation of the process causes dependencies in the sampled process measurements. A polymer quality sample is not drawn independently, but is dependent on the sampled (selected) reactor run. This sampling method is known as *multi-stage sampling*, and in this study, the design is two-stage: level-1 with “within-run” informa-

tion and level-2 with “between-run” information. This sampling structure implies that there is sampling variance at the reactor run level as well as sampling variance at the process measurements level. Hierarchical models are well suited to analysing data with a multi-stage design. The modeling hierarchy then describes the grouping of the reactor runs at level-2, and the measured process variables at level-1. At level-2, the group level, the *between-run* variation is described, while at level-1 the model describes the *within-run* variation.

### 3.2. Model development

To disentangle the information contained in the data, two sources of variation will be identified; within-reactor run variance and between-reactor run variance. Let  $Y_{ij}$  denote the the polymer quality of measurement  $i$  in reactor run  $j$ ,  $i = 1, \dots, n_j$ ,  $j = 1, \dots, J$ . The number of measurements  $n_j$  may vary from run to run. For example, for  $J = 10$  there are ten reactor runs. The observed between-reactor run variance is defined as the variance between the ten reactor run quality means. Within-reactor run variance is variation of measurements around their reactor run mean. In this case the within-reactor run variance will differ from reactor run to reactor run but a weighted average of the variances can express the within-reactor run variability for all ten reactor runs. The general idea is to identify the two sources of variation and to identify process variables that explain the variation within-reactor runs and to identify process variables that explain variation between-reactor runs.

In a more general approach attention is focused on an empty multilevel model, so called because it does not contain explanatory variables. This model is also known as the one-way random effects ANOVA model. In this model,  $Y_{ij}$  is considered to be the sum of a general mean  $\gamma_{00}$ , the overall mean of polymer quality measurements over reactor runs, a random reactor run effect  $U_{0j}$  (the deviation of the reactor run mean from the overall mean), and a random measurement effect  $R_{ij}$  (the deviation of the measurement from the reactor run mean), that is:

$$Y_{ij} = \gamma_{00} + U_{0j} + R_{ij} \quad (3)$$

with  $U_{0j} \sim N(0, \tau^2)$ , and  $R_{ij} \sim N(0, \sigma^2)$ . Further, it is assumed that

$$\begin{aligned} \text{cov}(U_{0j}, U_{0j'}) &= 0, \quad \forall j \neq j' \\ \text{cov}(U_{0j}, R_{ij'}) &= 0, \quad \forall j, j' \\ \text{cov}(R_{ij}, R_{i'j'}) &= 0, \quad \text{except for } i = i', \quad j = j'. \end{aligned}$$

Hence, measurements within the same reactor run are correlated, due to the fact that they share the same random component  $U_{0j}$ . This means that only quality measure-



ments within the same reactor run are correlated. It follows that:

$$\text{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \tau^2, & i \neq i', j = j' \\ 0, & j \neq j'. \end{cases}$$

The total variance of  $Y_{ij}$  can thus be written as:

$$\text{var}(Y_{ij}) = \text{var}(U_{0j}) + \text{var}(R_{ij}) = \tau^2 + \sigma^2 \quad (4)$$

which is the variance introduced at the level-2 of quality measurement means,  $\tau^2$ , the *between-reactor run variance*, plus the variance introduced at level-1 of quality measurements,  $\sigma^2$ , the *within-reactor run variance*. The partitioning of the total variance enables the computation of an *intra-class correlation coefficient*, denoted by  $\rho$ , that represents the proportion of variance that is accounted for by the grouping of quality measurements in reactor runs:

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2}. \quad (5)$$

Let  $\mathbf{x}_j = (x_{1j}, \dots, x_{Qj})$  denote the  $Q$ -dimensional vector with observed values of relevant process variables (e.g., stirrer power, feed flow rate, and the catalyst to feed flow ratio) all measurements in reactor run  $j$ . Typically, the first column of  $\mathbf{x}_j$  is a vector of ones for the intercept, while the other columns are variables (process measurements) that vary within a run. The level-1 regression model is given by:

$$Y_{ij} = x_{1ij}\beta_{0j} + x_{2ij}\beta_{1j} + \dots + x_{Qij}\beta_{(Q-1)j} + R_{ij}$$

which can be written in matrix notation as:

$$Y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}_j + R_{ij}, \quad (6)$$

where  $\boldsymbol{\beta}_j$  represents the  $Q \times 1$  vector of regression coefficients for run  $j$  and  $R_{ij} \sim N(0, \sigma^2)$  for  $j = 1, \dots, J$ .

Let  $\mathbf{z}_{qj} = z_{1qj}, \dots, z_{Kqj}$  denote the  $K$ -dimensional vector of covariates, with values of process variables like the thickness of the fouling layer at the wall of the reactor vessel, and the poisons concentration in a reactor run. They may influence the quality measurements across reactor runs. The level-2 regression model is given by:

$$\boldsymbol{\beta}_{qj} = z_{1qj}\gamma_{q0} + z_{2qj}\gamma_{q1} + \dots + z_{Kqj}\gamma_{q(K-1)} + U_{qj},$$

for  $q = 0, \dots, (Q-1)$  and  $j = 1, \dots, J$ . In matrix notation, it follows that:

$$\boldsymbol{\beta}_j = \mathbf{z}_j\boldsymbol{\gamma} + \mathbf{U}_j, \quad (7)$$

where  $\mathbf{U}_j \sim N(0, T)$  and  $\boldsymbol{\gamma}$  is the matrix of level-2 fixed regression coefficients. The diagonal elements of covariance matrix  $T$  represent the variation across reactor runs of random regression coefficients  $\boldsymbol{\beta}_j$ . The non-diagonal elements are not necessarily zero. That is, the random regression effects are allowed to correlate. Substitution of the level-2 model (7) into the level-1 model (6) yields the combined model:

$$Y_{ij} = \mathbf{x}_{ij}(\mathbf{z}_j\boldsymbol{\gamma} + \mathbf{U}_j) + R_{ij} = \mathbf{x}_{ij}\mathbf{z}_j\boldsymbol{\gamma} + \mathbf{x}_{ij}\mathbf{U}_j + R_{ij}. \quad (8)$$

From (8) it follows that the total variance of  $Y_{ij}$  is given by:

$$\begin{aligned} \text{var}(Y_{ij}) &= \text{var}(\mathbf{x}_{ij}\mathbf{z}_j\boldsymbol{\gamma} + \mathbf{x}_{ij}\mathbf{U}_j + R_{ij}) \\ &= \text{var}(\mathbf{x}_{ij}\mathbf{U}_j + R_{ij}) \\ &= \mathbf{x}_{ij}'\text{var}(\mathbf{U}_j)\mathbf{x}_{ij} + \text{var}(R_{ij}) \\ &= \mathbf{x}_{ij}'T\mathbf{x}_{ij} + \sigma^2, \end{aligned}$$

using the assumption that the residuals at different levels are uncorrelated. As a result, the two-level model presented in Eq. (8) allows for correlations of process measurements within each batch. Finally, the structure of the matrices are presented below which represents the unit-level representation for the entire sample. The level-1 model can be presented as

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{n_1 1} \\ \dots \\ \vdots \\ \dots \\ Y_{1J} \\ \vdots \\ Y_{n_J J} \end{bmatrix} = \begin{bmatrix} x_{111} & \dots & x_{Q11} & \vdots & \vdots \\ x_{1n_1 1} & \dots & x_{Qn_1 1} & \vdots & \vdots \\ \dots & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots & \dots \\ x_{11J} & \dots & x_{Q1J} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1n_J J} & \dots & x_{Qn_J J} & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{(Q-1)1} \\ \vdots \\ \beta_{0J} \\ \vdots \\ \beta_{(Q-1)J} \end{bmatrix} + \begin{bmatrix} R_{11} \\ \vdots \\ R_{n_1 1} \\ \vdots \\ R_{1J} \\ \vdots \\ R_{n_J J} \end{bmatrix}. \quad (9)$$

The level-2 model can be presented as

$$\begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{0J} \\ \dots \\ \vdots \\ \dots \\ \beta_{(Q-1)1} \\ \vdots \\ \beta_{(Q-1)J} \end{bmatrix} = \begin{bmatrix} z_{101} & \dots & z_{K01} \\ \vdots & \ddots & \vdots \\ z_{10J} & \dots & z_{K0J} \\ \dots & \dots & \dots \\ \vdots & & \ddots \\ \dots & \dots & \dots \\ z_{1(Q-1)1} & \dots & z_{K(Q-1)1} \\ \vdots & \ddots & \vdots \\ z_{1(Q-1)J} & \dots & z_{K(Q-1)J} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \vdots \\ \gamma_{0(K-1)} \\ \dots \\ \vdots \\ \dots \\ \gamma_{(Q-1)1} \\ \vdots \\ \gamma_{(Q-1)(K-1)} \end{bmatrix} + \begin{bmatrix} U_{01} \\ \vdots \\ U_{0J} \\ \dots \\ \vdots \\ \dots \\ U_{(Q-1)1} \\ \vdots \\ U_{(Q-1)J} \end{bmatrix} \quad (10)$$

It is possible to generalize this two-stage model to three or more levels. Interested readers, also in other generalizations, are referred to, e.g., Bryk and Raudenbush [16], Goldstein [17] or Snijders and Bosker [18].

#### 4. Estimation and testing

There are various approaches for estimating the parameters of a multilevel model. These include full maximum likelihood, restricted maximum likelihood, and Bayesian methods. These methods will give comparable estimates in large samples but somewhat different results can be obtained in small samples. Maximum likelihood methods are most often used for the multilevel model given in Eq. (8) where both error terms  $\mathbf{R}$  and  $\mathbf{U}$  are normally distributed.

Obtaining maximum likelihood estimates requires at least two steps since there is a no-closed expression for the maximizer of the likelihood, and an iterative scheme is required. The EM algorithm [19] and Fisher scoring [20] can be used to determine the estimates. The EM algorithm considers the problem of maximizing the likelihood as a problem in missing data. Substituting the level-2 model into the level-1 model yields the combined model,

$$\mathbf{Y}_j = \mathbf{x}_j \mathbf{z}_j \boldsymbol{\gamma} + \mathbf{x}_j \mathbf{U}_j + \mathbf{R}_j, \quad (11)$$

where  $\mathbf{R}_j \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_j})$  and  $\mathbf{U}_j \sim N(\mathbf{0}, \mathbf{T})$ . The random effects or level-2 error terms  $\mathbf{U}_j$  are considered as missing data. The so-called complete data that includes the missing data are  $(\mathbf{Y}_j, \mathbf{U}_j)$ . Then, full maximum likelihood estimates can be easily obtained if the missing data were observed. That is, an ordinary least squares (OLS) estimate can be computed for the fixed effects,  $\boldsymbol{\gamma}$ , and maximum likelihood estimates for  $\sigma^2$  and  $\mathbf{T}$ ,

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= \left( \sum_j \mathbf{z}_j' \mathbf{x}_j' \mathbf{x}_j \mathbf{z}_j \right)^{-1} \sum_j \mathbf{z}_j' \mathbf{x}_j' (\mathbf{y}_j - \mathbf{x}_j \hat{\mathbf{U}}_j) \\ \hat{\sigma}^2 &= J^{-1} \sum_j \hat{\mathbf{U}}_j' \hat{\mathbf{U}}_j \\ \hat{\mathbf{T}} &= N^{-1} \sum_j (\mathbf{y}_j - \mathbf{x}_j \mathbf{z}_j \hat{\boldsymbol{\gamma}} - \mathbf{x}_j \hat{\mathbf{U}}_j) (\mathbf{y}_j - \mathbf{x}_j \mathbf{z}_j \hat{\boldsymbol{\gamma}} - \mathbf{x}_j \hat{\mathbf{U}}_j)'. \end{aligned} \quad (12)$$

This completes the M-step of the EM-algorithm. From the standard normal distribution theory follows the conditional distribution of  $\mathbf{U}_j$  given  $\mathbf{y}_j$  from their joint normal distribution. This gives the conditional expectation of the missing data given the observed data, namely:

$$E(\mathbf{U}_j | \mathbf{y}_j, \boldsymbol{\gamma}, \sigma^2, \mathbf{T}) = (\mathbf{x}_j' \mathbf{x}_j + \sigma^2 \mathbf{T})^{-1} \mathbf{x}_j' (\mathbf{y}_j - \mathbf{x}_j \mathbf{z}_j \boldsymbol{\gamma}), \quad (13)$$

which defines the E-step of the EM-algorithm. The formulae for the restricted maximum likelihood estimates differ, and one indication of the difference is that restricted maximum likelihood estimates of the variance components take into account the loss of degrees of freedom resulting from the estimation of the fixed effects. However, the difference between the estimates is not expressible in simple algebraic form. Technical details can be found in, e.g., Bryk and Raudenbush [16] or Longford [21,22].

In some cases, estimates of the random regression coefficients  $\boldsymbol{\beta}_j$ , Eq. (10), are required. This multilevel model suggests two different estimates, one based on level-1 information, and one based on level-2 information. In a similar way to deriving the expected value of  $\mathbf{U}_j$ , an empirical Bayes estimate [23] of  $\boldsymbol{\beta}_j$  can be obtained that combines the two types of information. Formally, given estimates of the variance components and the fixed effects, the empirical Bayes estimate can be expressed as:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_j &= (\hat{\sigma}^{-2} \mathbf{x}_j' \mathbf{x}_j + \hat{\mathbf{T}}^{-1})^{-1} (\hat{\sigma}^{-2} \mathbf{x}_j' \mathbf{y}_j + \mathbf{T}^{-1} \mathbf{z}_j \hat{\boldsymbol{\gamma}}) \\ &= \boldsymbol{\Omega}_j \boldsymbol{\beta}_j^* + (\mathbf{I} - \boldsymbol{\Omega}_j) \mathbf{z}_j \hat{\boldsymbol{\gamma}}, \end{aligned} \quad (14)$$

where  $\beta_j^*$  is the OLS estimator for  $\beta_j$ , and

$$\Omega_j = \hat{T} \left( \hat{T} + \hat{\sigma}^2 (\mathbf{x}_j' \mathbf{x}_j)^{-1} \right)^{-1}. \quad (15)$$

In the present context, this shrinkage estimate of the random regression coefficients is constructed out of a weighted average of a within-reactor run regression estimate and a between-reactor run regression estimate. The weighting factor is defined as the ratio of the uncertainty regarding the regression estimate relative to the total variance of the quality measurements  $\mathbf{y}_j$ . The resulting estimates are said to be strengthened and shrink towards a regression plane defined by the set of process variables  $\mathbf{z}$ . It can be seen that high within-reactor run variance and/or a small number of level-1 quality measurements result in random regression estimates that are shrunk towards the overall mean since the within-reactor run information is unreliable. In the same way, the estimates of the random regression coefficients are mainly based on the within-reactor run information when the number of reactors is small and/or the between-reactor run variances are high.

As with ordinary linear regression, a  $t$ -test can be used to test hypothesis concerning a single fixed effect. A likelihood ratio or deviance test, defined as minus twice the log-likelihood, can be used to test several fixed parameters provided that full maximum likelihood estimates are used. The deviance can be interpreted as the lack of fit between model and data, and the larger the deviance, the poorer the fit to the data. In general, the deviance can be used to compare multilevel models with different fixed parts. The difference in deviance values, corresponding to two multilevel models fitted to the same data, is approximately  $\chi^2$  distributed with the difference in the number of fixed parameters as the

degrees of freedom. The deviance constructed out of the restricted maximum likelihood parameters can be used to compare models with different random parts but the same fixed part. Another approach is based on standard multivariate Wald tests (e.g., Snijders and Bosker [18]), which are approximately  $\chi^2$  distributed. They can be used to test the fixed part of the model but requires standard errors of the estimates and also the covariances among them.

## 5. Case study analysis

### 5.1. Analysing the data

Fig. 2 shows the distribution of polymer quality for the data of reactor 1 – grade A. It can be seen that the polymer quality produced differs considerably within and between the 20 reactor runs observed. The empty multilevel model (also called the one-way ANOVA with random effects), Eq. (3), provides useful information about the variation in the outcomes. In this case, data sets were available for both polymer grades (A and B) related to four different reactors. An empty model was fitted to each data set. All maximum likelihood estimates were obtained within three to six iterations using the EM algorithm. The estimation results are presented in Table 1.

The statistic to test formally whether the estimated variance in polymer quality across reactor runs is significantly greater than zero reduces to:

$$H = \sum_j n_j (\bar{Y}_j - \hat{\gamma}_{00})^2 / \hat{\sigma}^2 \quad (16)$$

which has a large sample  $\chi^2$  distribution with  $J - 1$  degrees of freedom under the null-hypothesis. Note that this

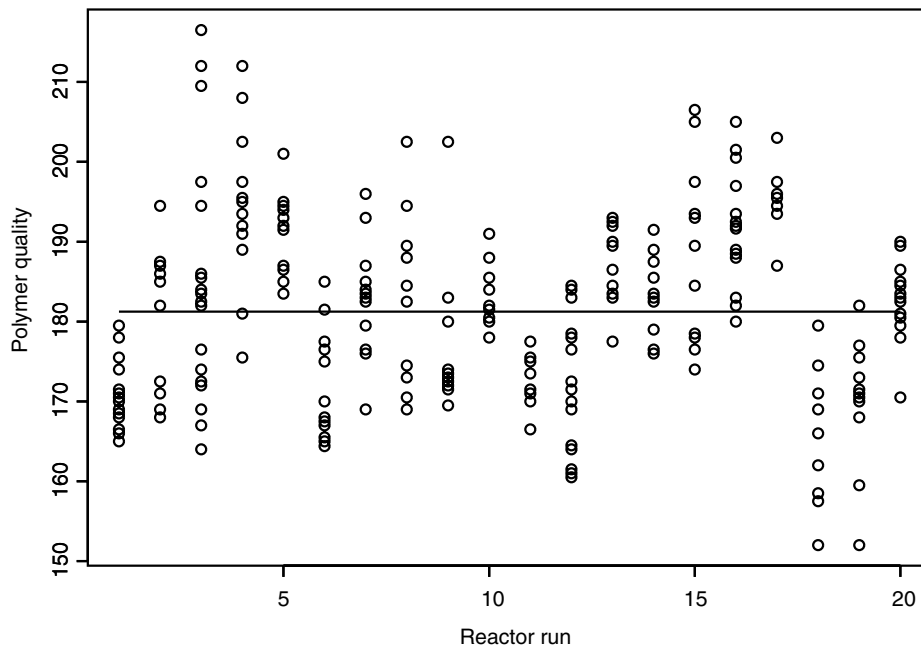


Fig. 2. Distribution of polymer quality for 20 runs of reactor 1, for production of polymer grade A. The horizontal line represents the general mean of the grade A polymer quality produced.



Table 1  
Estimated model parameters for the empty model, and corresponding intraclass correlation coefficient (ICC) estimates for the data of grade A and grade B

Grade – reactor	Parameter	Coefficient	se
A-1	$\gamma_{00}$	181.50	1.90
A-2	$\gamma_{00}$	175.50	1.60
A-3	$\gamma_{00}$	176.50	0.85
B-3	$\gamma_{00}$	231.50	3.35
B-4	$\gamma_{00}$	36.50	2.40
Grade – reactor	Parameter	Coefficient	ICC
A-1	$\tau^2$	67.30	.49
	$\sigma^2$	71.30	
A-2	$\tau^2$	53.30	.31
	$\sigma^2$	116.30	
A-3	$\tau^2$	18.00	.21
	$\sigma^2$	64.00	
B-3	$\tau^2$	222.30	.35
	$\sigma^2$	416.00	
B-4	$\tau^2$	75.30	.16
	$\sigma^2$	390.80	

statistic is just a weighted average of the squared differences between reactor run means and the general mean. In this case, all estimated variance components were significantly different from zero. As a result, for each grade and for each reactor, the produced polymer quality varied significantly both within and between runs. A 95% confidence interval for the grand mean  $\gamma_{00}$  equals:

$$\hat{\gamma}_{00} \pm 1.96\hat{\tau}, \quad (17)$$

and gauges the magnitude of the variation among reactor runs in their mean polymer quality per reactor. For the sample data of grade A of reactor 1, it is expected that 95% of the reactor run means fall within the range (165.43, 197.57). This indicates a substantial range in average polymer quality in this sample data. The other data sets show comparable ranges in average polymer quality. The intraclass correlation coefficient represents a quantification of the amount of variability between the reactor runs. For example, for the sample data of grade A of reactor 2, it follows that,

$$\hat{\rho}_{A2} = \hat{\tau}^2 / (\hat{\tau}^2 + \hat{\sigma}^2) = \frac{7.30}{7.30 + 10.80} = 0.31.$$

Therefore, around 31% of the total variance in polymer quality can be attributed to variance between reactor runs. For the other reactors and grades, the results are shown in the last column of Table 1. It can be concluded that a considerable amount of variance can be assigned to the variability between the reactor runs.

## 5.2. Several nested sources of heterogeneity

From Table 1, it can be seen that the estimated general mean,  $\gamma_{00}$ , of polymer quality varies across reactors and grades. Therefore, it can be expected that the observed process data are not only nested between runs (level-2), but

also between reactors (level-3) and between polymer grades (level-4). Thus, a multilevel model consisting of four levels is required to model the variation at different levels. Let  $Y_{ijkl}$  denote an observation of polymer quality  $i$  in reactor run  $j$ , and reactor  $k$  of grade  $l$ . The simplest multilevel model is fully unconditional and represents the variation in the outcome variable across the four levels (observations, reactor runs, reactors, grades), namely:

$$Y_{ijkl} = \gamma_{0000} + A_{000l} + B_{00kl} + U_{0jkl} + R_{ijkl}, \quad (18)$$

where  $R_{ijkl} \sim N(0, \sigma^2)$ ,  $U_{0jkl} \sim N(0, \tau^2)$ ,  $B_{00kl} \sim N(0, \phi^2)$ , and  $A_{000l} \sim N(0, \psi^2)$ . As a result, the total variability in the observed polymer quality is partitioned into four components. This representation allows the computation of variation within reactor runs, between reactors runs within reactors, and among reactors for different grades. Developing one model for this entire data set has the following three advantages:

- For analysis and process control, a model is desirable that generalizes over all reactors and polymer grades. This model can describe potential variations between reactors and grades. Making inferences across reactors and grades is much more complicated when using separate models.
- A model that generalizes over all reactors and grades is more robust than a model that is only valid for one reactor and/or one grade. However, this depends on to what extent the polymer processes in the different reactors/of different grades act in similar ways. A joint modeling approach of all measurements does not necessarily lead to a more robust model when there are no similarities between the polymer processes.
- An important argument for combining the sample data lies in the computational advantages. The random regression effects are estimated using shrinkage estimators, as discussed in Section 4. For example, a shrinkage estimate of the random regression effect  $B_{00kl}$  consists of composite estimates based on reactor means and estimates based on grade means of polymer quality, where the amount of shrinkage depends on the reliabilities of the reactor means and the grade means. These composite estimators are robust since they are based on different levels of information. The polymer processes in the different reactors/of different grades should act in a somewhat comparable way otherwise the information of other reactors/grades is not very useful for reactor/grade specific parameter estimation.

In the present case, at the polymer grade level only two polymer grades and at the reactor level only four different reactors were observed. With such a limited number of units at the grade and reactor level, it is not possible to consider them as samples from a population, but should rather be seen as selected grades and reactors. Therefore, the random effects  $A_{000l}$  and  $B_{00kl}$  defined in the four-stage multilevel model in Eq. (18), are considered to be fixed effects. In

this ANOVA approach, the reactors and the polymer grades are represented by dummy variables.

The measurements of polymer quality in each reactor run are regressed on the known predictors or covariates. It is assumed that these regressions are exchangeable (De Finetti [24]). This means that there is no information, other than the observed data, available to distinguish any of the random regression coefficients from any of the others. For example, in Eq. (7), the a priori assumptions about random regression coefficient  $\beta_j$  are independent of  $j$ . As a result, the same parametric structure for both polymer grades is assumed. That is, it is assumed that the mechanism of the process is similar for both grades. A justification for this assumption follows from process knowledge: the process is operated and controlled in the same way for both grades with the same variables, the difference between grades is due to the varying monomer feed ratios. Subsequently, the same parametric structure for the four reactors is assumed. A justification for this assumption is analogous to the previous assumption.

### 5.3. Model selection

Two influential level-1 explanatory variables were identified: the catalyst-to-feed-flow ratio ( $x_{1ij}$ ) and the feed flow rate ( $x_{2ij}$ ). These variables had the most explanatory power for polymer quality, and additional level-1 variables did not improve the proportion of explained variance by the model. The multilevel model with these predictor variables was also selected from all possible multilevel models given the level-1 variables via the deviance statistic. It is known that the operators adjust the values of the catalyst-to-feed-flow ratio and the feed flow rate in order to maintain a stable product quality. That is, the most influential explanatory process variables are the control variables of the operator. It can be expected that the effects of these variables on the polymer quality measurements fluctuate across reactor runs since, for instance, operators may act in different ways in controlling the process, and it is known that poisons influence the activity of the catalyst.

The matrix of level-1 covariates equals  $\mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij})$ . The level-1 model is given by:

$$Y_{ij} = \mathbf{x}_{ij}\beta_j + R_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, J \quad (19)$$

with  $R_{ij} \sim N(0, \sigma^2)$ . The random regression coefficients,  $\beta_j$ , are allowed to vary across reactor runs. Furthermore, the ANOVA approach is used to model variation in the regression coefficients across reactors and grades. This results in the following level-2 model:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}d_1 + \gamma_{02}d_2 + \gamma_{03}d_3 + \gamma_{04}d_4 + U_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}d_1 + \gamma_{12}d_2 + \gamma_{13}d_3 + \gamma_{14}d_4 + U_{1j}, \\ \beta_{2j} &= \gamma_{20} + \gamma_{21}d_1 + \gamma_{22}d_2 + \gamma_{23}d_3 + \gamma_{24}d_4 + U_{2j}, \end{aligned} \quad (20)$$

where  $\mathbf{U}_j = (U_{0j}, U_{1j}, U_{2j}) \sim N(\mathbf{0}, T)$ . The design matrix  $\mathbf{d} = (\mathbf{1}, \mathbf{d}_1, \dots, \mathbf{d}_4)$  consists of indicator variables  $d_1$  to  $d_4$ , that represent variation in polymer quality across the four

reactors, and variable  $d_4$  represents variation in polymer quality across grades. The design matrix  $\mathbf{d}$  is defined such that  $\gamma_{00}$  is the average intercept across level-2 reactor runs,  $\gamma_{01}$  the effect of reactor 1,  $\gamma_{02}$  the effect of reactor 2, and  $\gamma_{03}$  the effect of reactor 3. Subsequently, the sum of these parameters equals the effect of reactor 4. Fixed effect parameter  $\gamma_{04}$  represents the effect of polymer quality grade B. The variation in the other random regression effects are defined in the same way. It is not possible to incorporate interaction effects in the model since there are only two reactors producing grade A and grade B.

Eqs. (19) and (20) represent a full multilevel model, denoted as  $M_0$ . Deviance tests were performed to test assumptions of model  $M_0$ . That is, it was tested whether the observed polymer quality measurements varied across runs, reactors and grades. This led to the following reduced level-2 model:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{04}d_4 + U_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{14}d_4 + U_{1j} \\ \beta_{2j} &= \gamma_{20} + U_{2j}. \end{aligned} \quad (21)$$

The parameter estimates of model  $M_1$ , represented in Eq. (19) and (21) were obtained with restricted maximum likelihood. Convergence was achieved in 24 iterations, after which the log-likelihood reached stability at  $\log L = -3372$ . The parameter estimates are given in Table 2.

Model  $M_1$  allows the level-1 random regression coefficients to vary across reactor runs. Variations in polymer quality within a run are partly explained by the variations of the two control variables, the catalyst-to-feed-flow ratio and the feed-flow rate. The proportion of variance explained by these variables at level-1 can be computed by comparing the estimated variance in the empty model and the estimated variance based on the level-1 model in Eq. (19). Thus:

Proportion variance explained at level-1

$$= \frac{188.96 - 100.32}{188.96} \approx .47. \quad (22)$$

Table 2  
Estimated parameters for model  $M_1$

Fixed effect	Coefficient	se	t ratio	p value
$\gamma_{00}$	189.81	1.80	105.30	.00
$\gamma_{04}$	−9.49	1.70	−5.60	.00
$\gamma_{10}$	−8.72	.52	−16.80	.00
$\gamma_{14}$	6.95	.49	14.10	.00
$\gamma_{20}$	−.05	.03	−1.70	.09
Random effect	Variance	df	$\chi^2$	p value
$\tau_{00}^2$	207.73	113	694	.00
$\tau_{11}^2$	8.05	113	214	.00
$\tau_{22}^2$	.05	114	317	.00
$\sigma^2$	1100.32			

Adding the predictors at level-1 reduced the within-reactor run variance by 47%. It can be concluded that the variability in random regression effects across reactor runs cannot be explained by the different reactors. That is, the polymer quality produced does not vary significantly across reactors. Polymer measurements of grade B have significantly higher means, which follows from the fact that the mean polymer quality is supposed to differ between the two grades. Further, a notable feature is that  $\beta_{1j}$ , which are the within reactor run polymer quality – catalyst-to-feed-flow ratio slopes, are substantially less steep for polymer of grade A. This grade effect is probably the result from the different monomer feed ratios for both grades, which leads to differences in the reaction mechanism. As a result the sensitivity for the catalyst concentration can differ for both grades. This grade effect was not found in the polymer quality-feed flow rate. That is, the average value of  $\hat{\beta}_{2j}$ , for all  $j$  corresponding to grade A, do not differ significantly from the average value of  $\hat{\beta}_{2j'}$ , for all  $j'$  corresponding to grade B.

The relative large variance component for  $\beta_{1j}$  shows that there is a large variability in measurements due to changes in the catalyst to feed flow ratio between reactor runs. Given the sensitivity of the catalyst to poisons, this result might be an indication that, due to the recycle in the process, varying poison levels influence the activity of the catalyst. The variability in  $\beta_{1j}$  is shown in Fig. 3. The plot shows the regression lines for polymer quality against the catalyst to feed flow rate. Although the slope is negative for most lines, a large variation in slopes and intercepts can be observed.

Finally, the maximum likelihood point estimates for the covariance matrix  $T$  are:

$$\hat{T} = \begin{pmatrix} \hat{\tau}_{00}^2 & \hat{\tau}_{01}^2 & \hat{\tau}_{02}^2 \\ \hat{\tau}_{10}^2 & \hat{\tau}_{11}^2 & \hat{\tau}_{12}^2 \\ \hat{\tau}_{20}^2 & \hat{\tau}_{21}^2 & \hat{\tau}_{22}^2 \end{pmatrix} = \begin{pmatrix} 207.73 & -15.60 & 2.00 \\ -15.60 & 8.05 & -.20 \\ 2.00 & -.20 & .05 \end{pmatrix}. \quad (23)$$

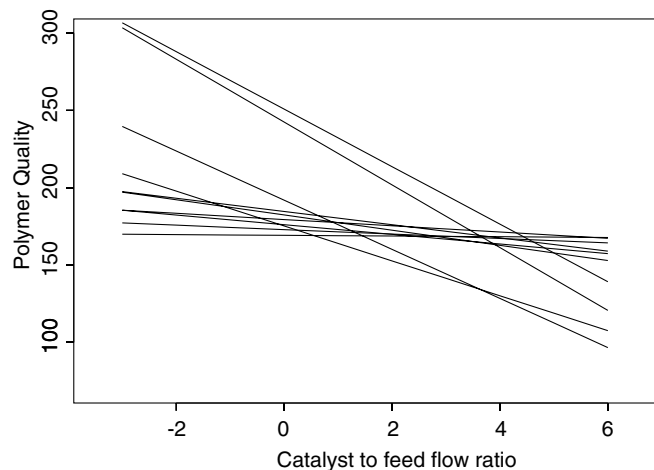


Fig. 3. Regression lines of polymer quality against catalyst to feed flow ratio (mean centered) for a random sample of 10 reactor runs.

There is a negative correlation between the random regression coefficients  $\beta_{1j}$  and  $\beta_{2j}$  of 32%. This means that relatively high polymer quality-catalyst to feed flow ratio slopes within a reactor run correspond to relatively high negative polymer quality-feed flow rate slopes. Here, interest was focused on maintaining a stable product quality, and the feed flow rate was used to control the effect of the catalyst to feed flow ratio. That is, when the reaction rate increases, an adjustment of the feed flow was necessary to control the reactor temperature.

#### 5.4. Model assessment

Inferences based on the multilevel model depend on their validity on the justifiability of the assumptions of the model. Assumptions are made at each level of the model, and misspecification at one level can affect results at other levels. The multilevel model with two levels, Eq. (19) and (21), contains several assumptions. It is assumed that level-1 residuals are identical, independent and normally distributed, and the level-2 residual vectors  $U_j$  are identical multivariate normally distributed, and that they are independent among the  $J$  reactor runs. The residuals at level-1 and level-2 are assumed to be independent. The predictors at each level are independent of the residuals at level-1 and level-2.

Most of the diagnostic tools used for detecting inadequacies within a linear regression model can also be used for checking the adequacy of a multilevel model. In this paper, only a number of diagnostics are mentioned, see, for a complete overview, e.g., Bryk and Raudenbush [16], Goldstein [17], Snijders and Bosker [18]. The normality assumption of the residuals at level-1 was checked by computing separate normal probability plots, in which each estimated residual value is plotted against its expected value under normality, for each reactor run. Only some of the plots suggested that the error distribution was not normal. In general, if the normality assumption at level-1 fails, it will introduce bias into standard errors at both levels that influences the computation of confidence intervals and hypothesis tests. Failure of normality at level-2 affects the confidence intervals and hypothesis tests for the fixed effects at level-2. Checking for normality at level-2 is more complicated since the random effects are not directly observed. Further, often a relatively small number of level-2 units are observed and tests for normality, like the chi-square test, and the Kolmogorov–Smirnov test are based on sufficiently large samples. Here, only 116 observations at level-2 were observed which makes it difficult to verify the assumption of normality at level-2.

The rate of convergence of the EM-algorithm can be considered as a diagnostic tool since rapid convergence indicates that the data are highly informative and slow convergence indicates numerical difficulties and/or that the data are sparse which may suggest that one or more random coefficients should be modeled differently. Further, an analysis of the correlations among level-1 residuals

can be informative about the number of random effects. That is, one or more random effects can be constrained to be zero when a high degree of multicollinearity is found. Finally, plots of residuals against predictors can identify relationships between them. These formal and informal diagnostics were performed for the fitted multilevel models given the polymer quality data. None of the diagnostics suggested a serious violation of one of the model assumptions.

A goodness-of-fit measure is the explained variance  $R^2$  that summarizes the fit of the regression by the proportion of variance explained. Gelman and Pardoe [25] developed an  $R^2$  based on the proportion of variance at each level of the multilevel model. The proportion of variance explained at level-1 is defined by:

$$R_1^2 = 1 - \frac{\frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \mathbf{x}_{ij} \hat{\boldsymbol{\beta}}_j)^2}{\frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2}, \quad (24)$$

and at level-2:

$$R_2^2 = 1 - \frac{\frac{1}{J} \sum_{j=1}^J (\hat{\boldsymbol{\beta}}_j - \mathbf{D} \hat{\boldsymbol{\gamma}}_j)^t (\hat{\boldsymbol{\beta}}_j - \mathbf{D} \hat{\boldsymbol{\gamma}}_j)}{\frac{1}{J} \sum_{j=1}^J (\bar{y}_j - \bar{y})^2}, \quad (25)$$

where design matrix  $\mathbf{D}$  represents the structure as defined in Eq. (21). The  $R^2$  is close to zero when the estimated residual variance in the numerator is approximately equal to the estimated variance of the denominator, the variance of the data. Thus, the  $R^2$  in Eq. (24) and (25), can be considered to summarize the explained variance of the regression within reactor runs and of the regression across reactor runs.

Here, this method can be generalized by splitting the total proportion of variance explained by the polymer quality grade. Let subscript  $g$  denote the polymer quality grade,  $J_g$  the number of batches in grade  $g$ , and  $\bar{y}_g$  the mean polymer quality of grade  $g$  in the sample. Then  $R_{1,g}^2$  and  $R_{2,g}^2$  respectively equal:

$$R_{1,g}^2 = 1 - \frac{\frac{1}{N_g} \sum_{j \in g} (\mathbf{y}_j - \mathbf{x} \hat{\boldsymbol{\beta}}_j)^t (\mathbf{y}_j - \mathbf{x} \hat{\boldsymbol{\beta}}_j)}{\frac{1}{N_g} \sum_{j \in g} (\mathbf{y}_j - \bar{y}_g)^t (\mathbf{y}_j - \bar{y}_g)} \quad (26)$$

$$R_{2,g}^2 = 1 - \frac{\frac{1}{J_g} \sum_{j \in g} (\hat{\boldsymbol{\beta}}_j - \mathbf{D} \hat{\boldsymbol{\gamma}}_j)^t (\hat{\boldsymbol{\beta}}_j - \mathbf{D} \hat{\boldsymbol{\gamma}}_j)}{\frac{1}{J_g} \sum_{j \in g} (\bar{y}_j - \bar{y}_g)^2}.$$

As a result, the proportion of explained variance at level 1 is obtained by calculating  $R_{1,A}^2$  and  $R_{1,B}^2$ . For multilevel model  $M_1$  the amount of explained variance obtained is  $R_{1,A}^2 = 0.62$  and  $R_{1,B}^2 = 0.70$ . At level-2, the amount of explained variance obtained is  $R_{2,A}^2 = 0.98$  and  $R_{2,B}^2 = 0.95$ .

Finally, note that the proportion of variance explained by both grades, as defined in Eq. (24) and (25) overestimate the true amount of explained variance since the reactor run polymer quality means of grade A and grade B substantially differ. Consider  $\hat{y}_A$ , and  $\hat{y}_B$  to be predictions under the multilevel model for grade A and B, respectively. Further, let  $\bar{y}_A \neq \bar{y}_B$  and let  $\bar{y}_B > \bar{y}_A$ . Then,  $\bar{y}_A < \bar{y} < \bar{y}_B$ , and it follows that for any grade  $g$  that

$\sum_{j \in g} (\mathbf{y}_j - \hat{y}_g)^t (\mathbf{y}_j - \hat{y}_g) < \sum_{j \in g} (\mathbf{y}_j - \bar{y})^t (\mathbf{y}_j - \bar{y})$ . As a result, the denominator of  $R_1^2$ , in Eq. (24), represents the variance in the data which is much larger than the residual variance defined in the numerator due to substantial differences in polymer grade means.

### 5.5. Prediction of future observations

After carrying out a multilevel analysis, prediction of polymer qualities given level-1 and level-2 characteristics is considered. Let,  $\tilde{\mathbf{y}}_j$  denote an unobserved measurement in the  $j$ th reactor run, where reactor run  $j$  is not necessarily in the observed sample of measurements. A well-known multilevel prediction rule is given by,

$$\tilde{y}_j = \tilde{\mathbf{x}}_j \hat{\boldsymbol{\beta}}_j, \quad (27)$$

where  $\hat{\boldsymbol{\beta}}_j$  is the empirical Bayes estimate in Eq. (14), and  $\tilde{\mathbf{x}}_j$  the level-1 characteristics corresponding to the unobserved measurement from a reactor run in the sample. Afshartous and De Leeuw [26] showed some analytical results for this multilevel prediction rule. They showed that the estimator in Eq. (27) minimizes the mean square error, namely:

$$E(\tilde{y}_j | \mathbf{x}, \mathbf{Y}) = \tilde{\mathbf{x}}_j \hat{\boldsymbol{\beta}}_j. \quad (28)$$

The data set was split up into an estimation set and a validation set, by selecting every final measurement of each reactor run for the validation set. The parameters of model  $M_1$  were estimated given the estimation data. Subsequently, the estimates of  $\hat{\boldsymbol{\beta}}_j$  were used to predict the new observations  $\tilde{y}_j$  for every run  $j$ . The results are shown in Fig. 4. It can be seen that predictions of high polymer qualities show more variation and are less close to the true observed values. However, after recalculating the  $R^2$  for both grades, the level-1 fit-statistic showed just a small decrease for both grades with  $R_{1,A}^2 = .60$  and  $R_{1,B}^2 = .68$ .

Predicting the outcome of a polymer quality,  $\tilde{y}_{j'}$  ( $j' \neq j$  for  $j = 1, \dots, J$ ) from an unobserved reactor run is more difficult. Then, the reactor run specific random

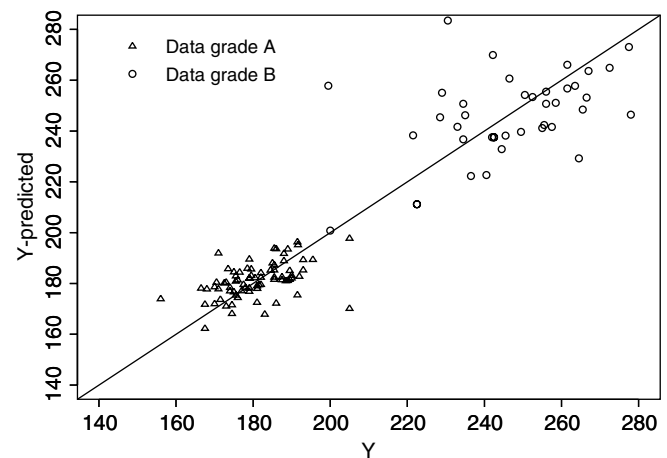


Fig. 4. Predicting a new observation for every reactor run. Measured versus model predicted values.



coefficients are unknown and are to be predicted at level-2 given characteristics  $\tilde{\mathbf{z}}_{j'}$  of the unobserved reactor run  $j'$ . That is,

$$\tilde{y}_{j'} = \tilde{\mathbf{x}}_{j'} \tilde{\mathbf{z}}_{j'} \hat{\gamma}. \quad (29)$$

In the present case, the limited amount of level-2 information causes the predictions to shrink towards the polymer quality grade mean. Within this approach, the predictions can only be improved by useful level-2 variables. Another approach is to observe a new reactor run  $j'$  for a certain time  $t_s$  such that  $n_{t_s}$  quality samples are obtained. The multilevel model parameters are estimated given the sample including a relatively small sample of reactor run  $j'$ . Then, random regression coefficient estimates are obtained for all reactor runs, including  $j=j'$ . Due to the advantage of shrinkage estimators, and the advantage of borrowing strength, the random regression estimates are stable and will improve the multilevel predictions. In an online application, this technique can be applied iteratively. Every time a new polymer quality sample of the reactor run is obtained, the model can be updated to improve its performance.

## 6. Comparison with alternative models

One alternative model is a fixed effects model for the entire sample of measurements with dummy variables for grades and reactors. In this ANOVA model, variation in the measurements across reactors and grades are recognized but the clustering of measurements within a reactor

run, and run-specific covariates are not included. This model  $M_2$  is given by,

$$\begin{aligned} \mathbf{Y} &= \mathbf{x}\boldsymbol{\beta} + \mathbf{e} \\ \boldsymbol{\beta} &= \mathbf{d}\boldsymbol{\gamma}, \end{aligned} \quad (30)$$

where  $\mathbf{d}$  represents the design matrix. Note that the regression coefficients are considered to be fixed effects and are not allowed to vary randomly across reactor runs. That is, the equation for the regression coefficients does not include a random error term as in model  $M_0$ . The residuals,  $\mathbf{e}$ , are independent and normally distributed. To illustrate the modeling error when the within reactor run dependencies between measurements are ignored, the parameters of this fixed effects model, Eq. (30), are estimated given the sample data. In Fig. 5 the model fitted versus the measured values are plotted for model  $M_1$  (on the left) and  $M_2$  (on the right). A significant difference in model performance is observed when comparing the left plot with the right plot of Fig. 5. The explanatory capability of model  $M_2$  is considerably weaker than that of model  $M_1$ . This result is in agreement with the calculated intraclass correlations coefficients (see Table 1) that indicate a high proportion of explained variance at the level of reactor runs.

An ANOVA approach for the reactor runs in model  $M_2$ , analogous to that for the four reactors, is not possible since this would require  $115 \times 3$  extra model parameters, which would result in highly unstable parameter estimates.

A second alternative is to fit separate regression models within each run. However, estimates of run-specific coefficients can be very imprecise or even impossible to obtain, and the standard deviations would be much larger than

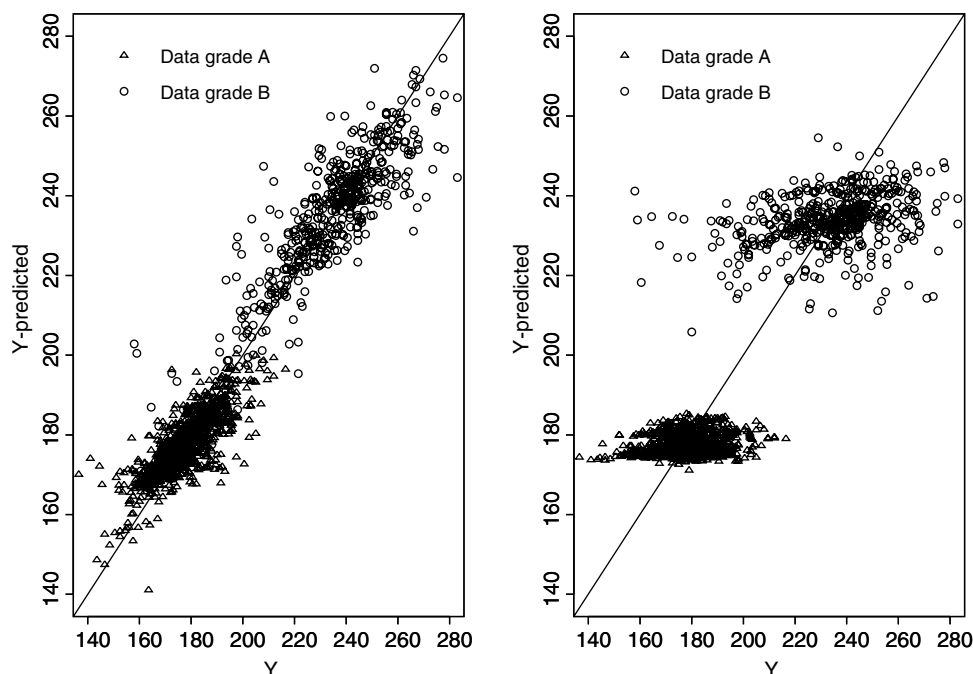


Fig. 5. Fitted values based on model  $M_1$  (left) and based on model  $M_2$  (right) plotted against the measured values of polymer quality.



those from a multilevel model due to the absence of pooling information across runs.

## 7. Discussion

The performed analysis showed that the semi-continuous operation mode leads to dependencies in the data collected. It was shown that when ignoring the grouping structure of the data, it is not possible to obtain a robust model for multiple reactor runs. However, by applying a hierarchical modeling approach, the proportion of explained variance in polymer quality improves considerably. With the model developed, important factors which explain polymer quality variance have been identified. From the measured process variables there are two control variables of the process: the feed-flow rate and the catalyst-to-feed-flow ratio.

The model also provides opportunities for process analysis, from which useful information for operation can be obtained. It is important to note that, when grouping of the process data is significant (indicated by high intraclass correlation) ignoring the grouping structure can lead to serious modeling errors and thereby to false interpretations of operation of the process. Therefore, it is argued that an analysis of batch-wise and semi-continuous operated processes data should take account of the hierarchical structure of the data.

Nonetheless there still remain some possible developments in this research for the future. First of all, it was identified that it is difficult to validate the assumption of normality of the model. The ideal solution to this problem would be obtaining considerably more data. However, in most cases this is not possible due to the characteristics of the process. An alternative could be to investigate the use of a *t*-distribution at level-1 or/and level-2. Seltzer [27] showed that a *t*-distribution provides robust estimates of the fixed effects that are less sensitive to outlier values. Secondly, there is still a proportion of unexplained variance by the model. Considering the process and its time constants, it is possible that there might be a moderate autocorrelation between subsequent polymer quality samples. Improvement of explained variance by the model could be achieved by applying an autoregressive model structure. However, a problem with estimating autoregressive multilevel models is that such models are complex and, as a result, it is yet not possible to estimate them using existing standard software.

## References

- [1] C. Kiparissides, Polymerization reactor modeling: a review of recent developments and future directions, *Chem. Engng. Sci.* 51 (1996) 1637–1659.
- [2] C. Kiparissides, J. Morris, Intelligent manufacturing of polymers, *Comp. Chem. Engng.* 20 (1996) S1113–S1118.
- [3] M. Ohshima, M. Tanigaki, Quality control of polymer production processes, *J. Process Contr.* 10 (2000) 135–148.
- [4] P. Terwiesch, M. Agarwal, D.W.T. Rippin, Batch unit optimization with imperfect modelling: a survey, *J. Process Contr.* 4 (1994) 238–258.
- [5] P. Terwiesch, D. Ravemark, B. Schenker, D.W.T. Rippin, Semi-batch process optimization under uncertainty: theory and experiments, *Comp. Chem. Engng.* 22 (1998) 201–213.
- [6] M. Aitkin, N. Longford, Statistical modelling issues in school effectiveness studies, *J. Royal Stat. Soc., Series A* 149 (1986) 1–43.
- [7] P. Nomikos, J.F. MacGregor, Monitoring of batch processes using multi-way principal component analysis, *AIChE J.* 40 (1994) 1361–1375.
- [8] B. Li, J. Morris, E.B. Martin, Model selection for partial least squares regression, *Chemometr. Intell. Lab. Syst.* 64 (2002) 79–89.
- [9] P. Nomikos, J.F. MacGregor, Multi-way partial least squares in monitoring batch processes, *Chemometr. Intell. Lab. Syst.* 30 (1995) 97–108.
- [10] P. Nomikos, J.F. MacGregor, Multivariate SPC charts for monitoring batch processes, *Technometrics* 37 (1995) 41–59.
- [11] A. Kassidas, J.F. MacGregor, P.A. Taylor, Synchronization of batch trajectories using dynamic time warping, *AIChE J.* 44 (1998) 864–875.
- [12] H.J. Ramaker, E.N.M. van Sprang, J.A. Westerhuis, A.K. Smilde, Dynamic time warping of spectroscopic batch data, *Anal. Chim. Acta* 498 (2003) 133–153.
- [13] K.S. Whiteley, T.G. Heggs, H. Koch, R.L. Mawer, W. Immel, Polyolefins, in: *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH publishers, New York, 2000.
- [14] B. Roffel, B.H.L. Betlem, *Physical and Empirical Process Modelling for Prediction and Control*, Wiley, New York, 2006.
- [15] K.R. Westerterp, W.P.M. van Swaaij, A.A.C.M. Beenackers, *Chemical Reactor Design and Operation*, Student ed., Wiley, New York, 1987.
- [16] A.S. Bryk, S.W. Raudenbush, *Hierarchical Linear Models*, Sage Publications Inc., Newbury Park, 1992.
- [17] H. Goldstein, *Multilevel Statistical Models*, third ed., Oxford University Press, New York, 2003.
- [18] T.A.B. Snijders, R.J. Bosker, *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*, Sage Publishers, London, 1999.
- [19] A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. Royal Stat. Soc., Series B* 39 (1977) 18.
- [20] N.T. Longford, A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects, *Biometrika* 74 (1987) 817–827.
- [21] N.T. Longford, *Random Coefficient Models*, Clarendon Press, Oxford, 1993.
- [22] N.T. Longford, Hierarchical-models social-sciences, *J. Educ. Behav. Stat.* 20 (1995) 205–209.
- [23] C.N. Morris, Parametric empirical Bayes inference: theory and applications, *J. Am. Stat. Assoc.* 78 (1983) 47–65.
- [24] B. De Finetti, Foresight, its logical laws, its subjective sources, in: H.E. Kyburg Jr., H.E. Smokler (Eds.), *Studies in Subjective Probability*, Wiley, New York, 1964, pp. 93–158.
- [25] A. Gelman, I. Pardoe, Bayesian measures of explained variance and pooling in multilevel (hierarchical) models, *Technometrics* 48 (2006) 241–251.
- [26] D. Afshartous, J. de Leeuw, Prediction in multilevel models, *J. Educ. Behav. Stat.* 30 (2005) 109–139.
- [27] M.H. Seltzer, Sensitivity analysis for fixed effects in the hierarchical model: a Gibbs sampling approach, *J. Educ. Stat.* 18 (1993) 207–235.